D.A. Johnston
Manned

1 up

TEXAS A&I UNIVERSITY

TEXAS A&I UNIVERSITY KINGSVILLE TEXAS 1925

Kingsville, Texas

(NASA-CR-129507) [INVESTIGATION OF ESTIMATORS OF PROBABILITY DENSITY FUNCTIONS] Final Report F.M. Speed (Texas A&I Univ., Kingsville.) Jul. 1972 27 p CSCL 09B G3/08

N73-12178

Unclas 16368

27

TEXAS A & I UNIVERSITY

RESEARCH GRANT #NGR44-073-003

TECHNICAL REPORT #5

FINAL REPORT

F. M. Speed

July, 1972

I

DEPARTMENT OF

MATHEMATICS

SUMMARY

This report is the final report for the research carried out under the NASA Grant #NGR 44-073-003. This report contains parts of the abstracts or introductions for the four technical reports submitted under this grant. All of the research has been reported in these reports. In addition, a number of graphs have been included with this report that have not been submitted before. The analysis and interpretation of the graphs can be found in Technical Report No. 4.

TECHNICAL REPORT NO. 1

THE GENERATION OF RANDOM NUMBERS ON THE IBM 360/44

In recent years there has been a need for "reliable" random numbers which have a uniform distribution over the unit interval U(0,1). At the present time there are three modes of providing random numbers for use on digital computers, specifically the IBM 360/44: external provision, internal generation by a random process, and internal generation of sequences of digits by a recurrence relation. The most common mode in use today is the internal generation of sequences of digits by a recurrence relation.

The use of random numbers in important large scale simulation programs, often referred to as Monte Carlo applications, has increased the need for good quality random numbers.

In the last few years there has been speculation from noted authorities in the field of random numbers that one of the established random number generators employed on many IBM 360 computers is inadequate for scientific results.

This paper is a summary of the generation and testing of random numbers on the IBM 360/44 computer.

TECHNICAL REPORT NO. 2

COMPUTER PROGRAMS AND DOCUMENTATION

This report contains a description of the various statistical tests that were used to check out random number generators. The tests contained in this report are by no means all the possible tests that can be run. A total of 12 different tests were considered. And from these, 6 were choosen to be used. Among those not included in this report are such tests as the poker test, the coupon test, the spectral test, and so on. The 6 tests that were choosen were done so becuase of the properties that they appeared to exhibit. Also, these are the most classical tests that are run. One test, which was not included, is the spectral test. The reason why it was not included was because of the need of a very large computer. If such a computer would have been available, we would have included this test because it is a very powerful test.

The tests included in this report are the frequency test, the max t test, the run test, the lag product test, the gap test, and the matrix test. This report is divided into three major sections. The first section concerns those tests of goodness of fit; and under this we have the frequency and the max t test. The next section consists of those tests of independence; and this includes the run, the lag product, the gap, and the matrix test. The final section gives documentation on the use of these various tests as well as a listing of the programs.

The discussion in parts 2 and 3 makes the following assumptions. We have a sequence U_1, U_2, U_3, and so that come from a pseudorandom number generator that is supposed to be generating random numbers from a uniform distribution and the numbers are supposed to be independently distributed. For the remainder of this report the terminology "random numbers" will be used to mean pseudorandom numbers.

TECHNICAL REPORT NO. 3

INVESTIGATION OF THE SPECHT DENSITY ESTIMATOR

This preliminary study shows that the Specht estimator is highly dependent upon the choice of N and S. As N gets large, it is still necessary to choose s with care. A too small choice of S will produce very bad results. While the Specht estimator does a fair job of estimating the normal distribution, it is poor for estimating the uniform density. A more detailed study is being done to compare

this estimator with other estimators found in the literature.

TECHNICAL REPORT NO. 4

INVESTIGATION OF ESTIMATORS OF PROBABILITY DENSITY FUNCTIONS

The use of density functions may be very useful in analyzing large amounts of data. This data, which is often collected by remote sensors, is obtained in large quantities. Manual analysis of this data is impractical, thus an automated procedure would be very helpful in handling the data.

The collection and analyzing of remote sensing data can be helpful in many areas. For example, a source of water pollution can be located by identifying the slainity of the water in rivers, lakes and oceans. Also, agricultural crops, and many disease plants, can be identified by the use of proper data. On land, geology may be aided by the possible detection of mineral deposits. These are but a few of the areas that can be helped by collecting and analyzing remote sensing data.

To understand the whole problem, let us consider the problem of identifying agriculture crops. An observation vector from a scanner is obtained and this vector, X, will be used to identify the crop.

If the density function of each crop were known, then the probability that X came from a certain density function could be calculated. However, the density functions are not

known. Thus, an attempt is made to estimate density functions.

In this paper, we used Parzen's estimators of the form:

$$f_n(x) = \frac{1}{n \cdot s(n)} \sum_{j=1}^{n} K\left(\frac{x - x_j}{s(n)}\right)$$

where n is the number of points in the sample, x_j is one of the points in the sample, s(n) is the smoothing parameter and K(y) is one of the following functions:

$$K(y) = (2\pi)^{-1/2} e^{-y^2/2}$$

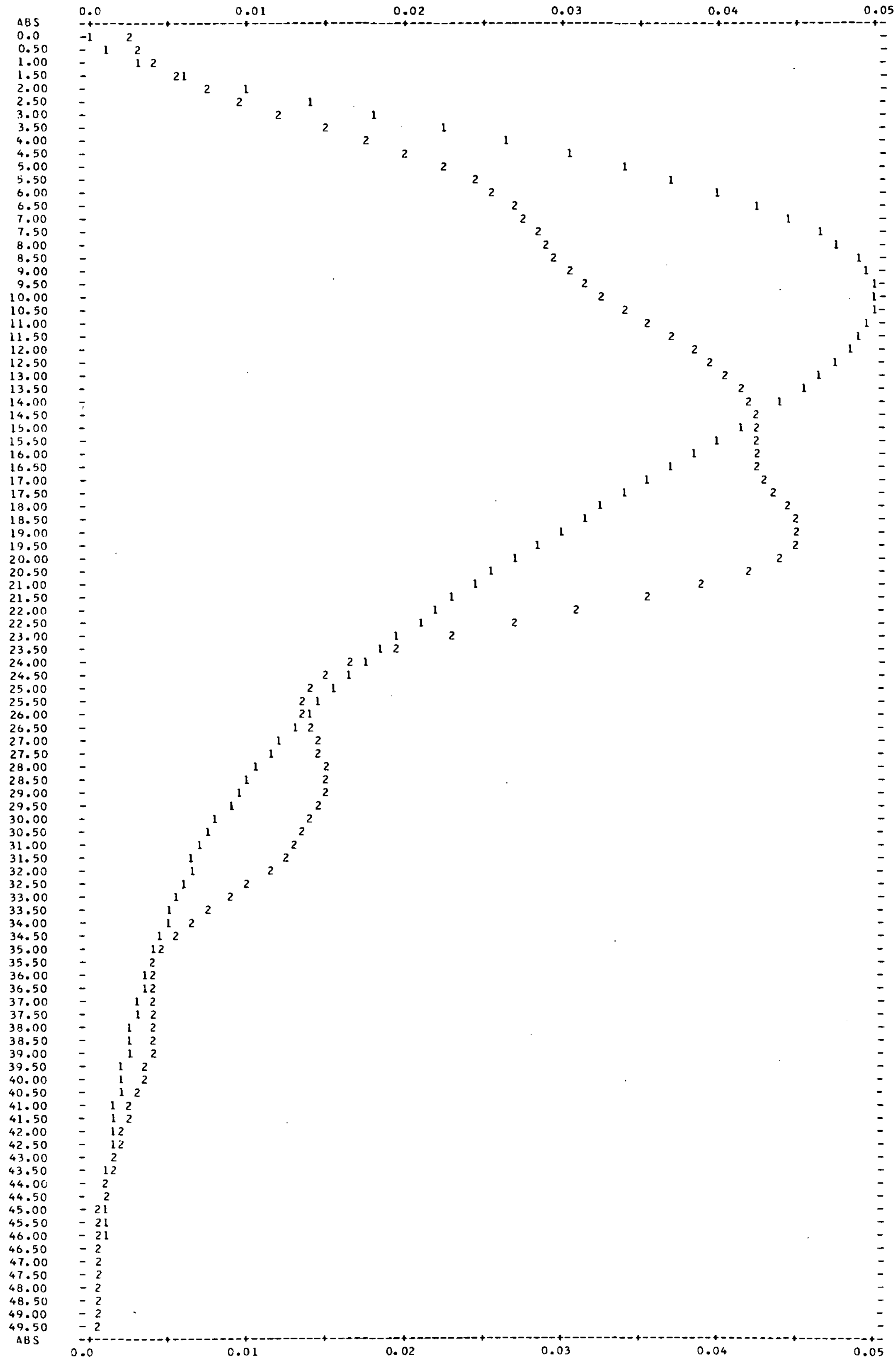
$$K(y) = (2\pi)^{-1} (\sin(y/2)/(y/2))^2$$

$$K(y) = \begin{cases} 0 & \text{if } |y| \geq 1 \\ 1 - |y| & \text{if } |y| < 1. \end{cases}$$

We want to determine if the estimation of density functions is indeed practical. Also, if the K function and the choice of the smoothing parameter s, are critical in estimating a density function.

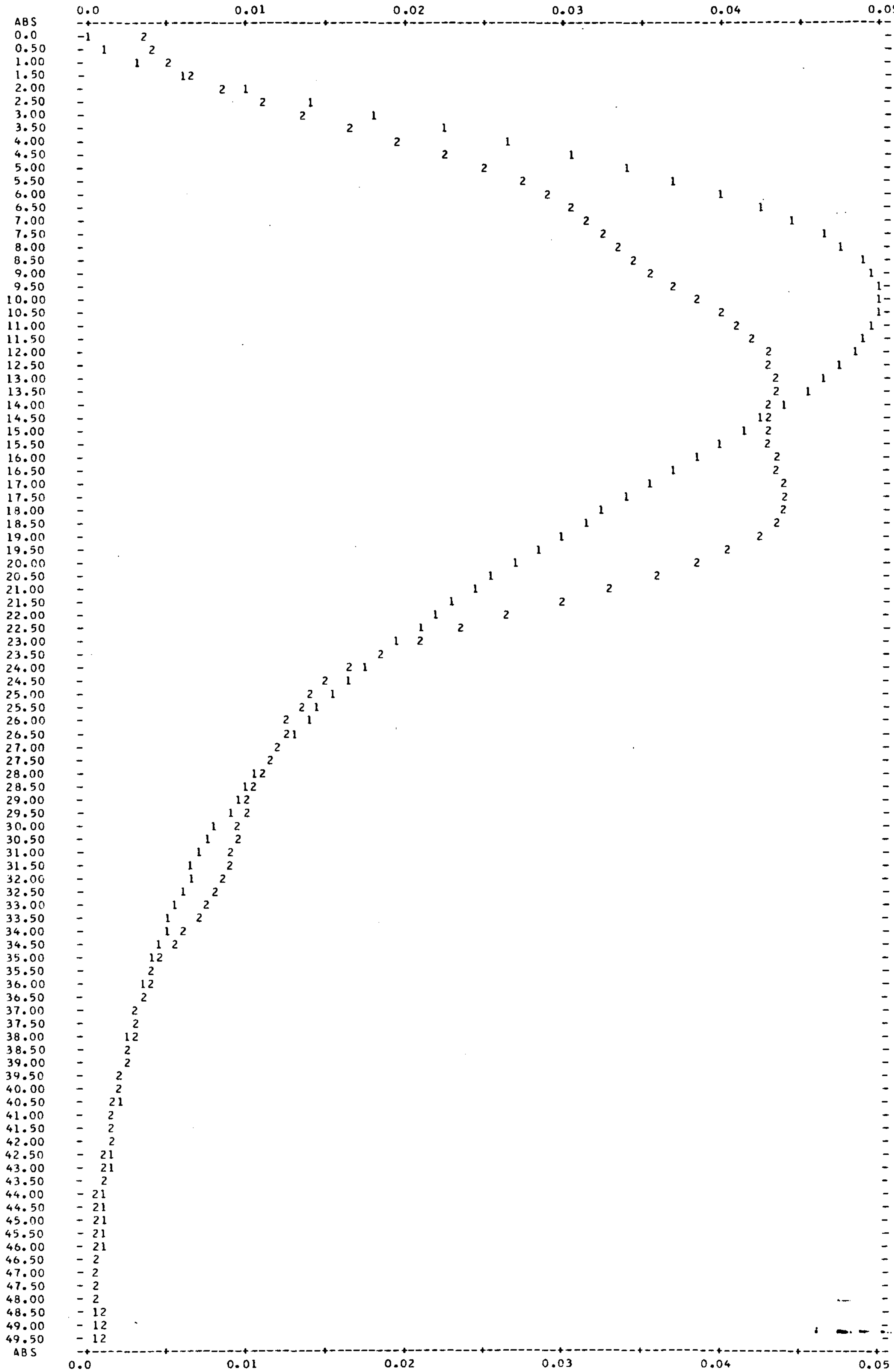
GRAPH

PLOTS OF TRUE AND ESTIMATED DENSITY FUNCTIONS
1 DENOTES TRUE G(3,.2), 2 DENOTES ESTIMATED
K(Y) = (1/(2*PI))*(SIN(Y/2)/(Y/2))**2
N = 100 S = 0.80



GRAPH

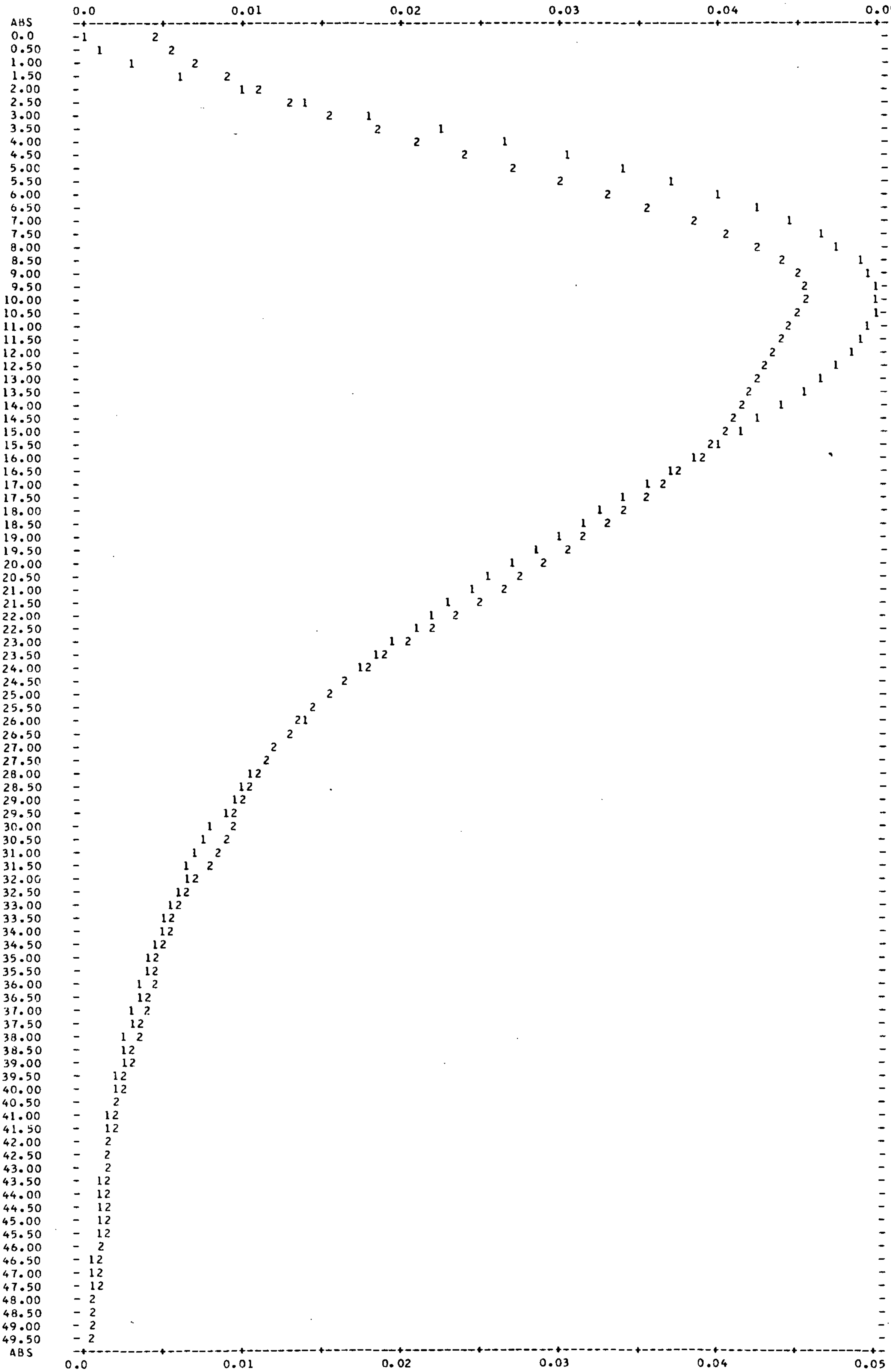
PLOTS OF TRUE AND ESTIMATED DENSITY FUNCTIONS
1 DENOTES TRUE G(3,.2), 2 DENOTES ESTIMATED
K(Y) = (1/(2*PI))*(SIN(Y/2)/(Y/2))**2
N = 200 S = 0.80



7

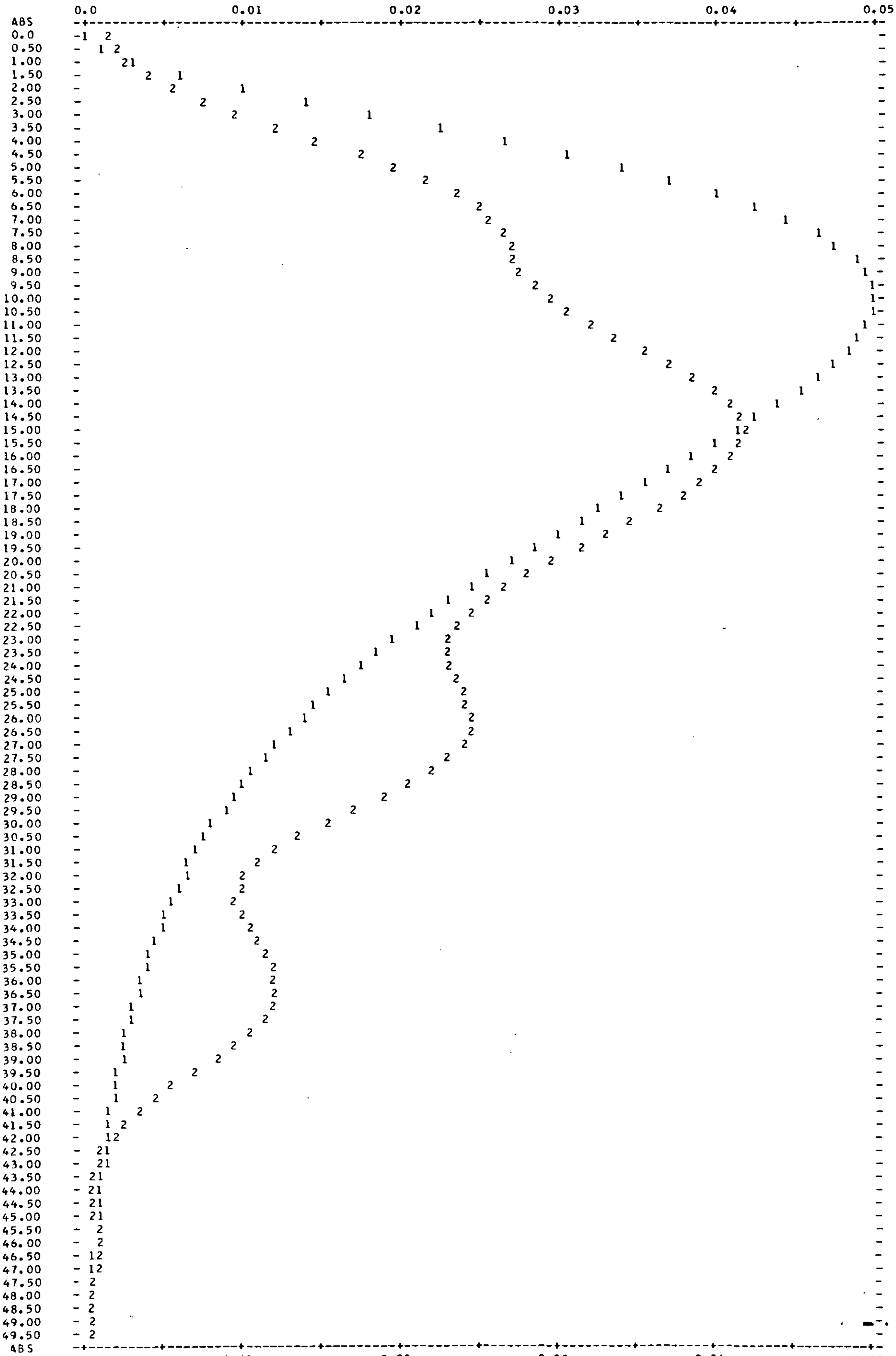
GRAPH

PLOTS OF TRUE AND ESTIMATED DENSITY FUNCTIONS
1 DENOTES TRUE G(3,.2), 2 DENOTES ESTIMATED
K(Y) = (1/(2*PI))*(SIN(Y/2)/(Y/2))**2
N = 1000 S = 0.80



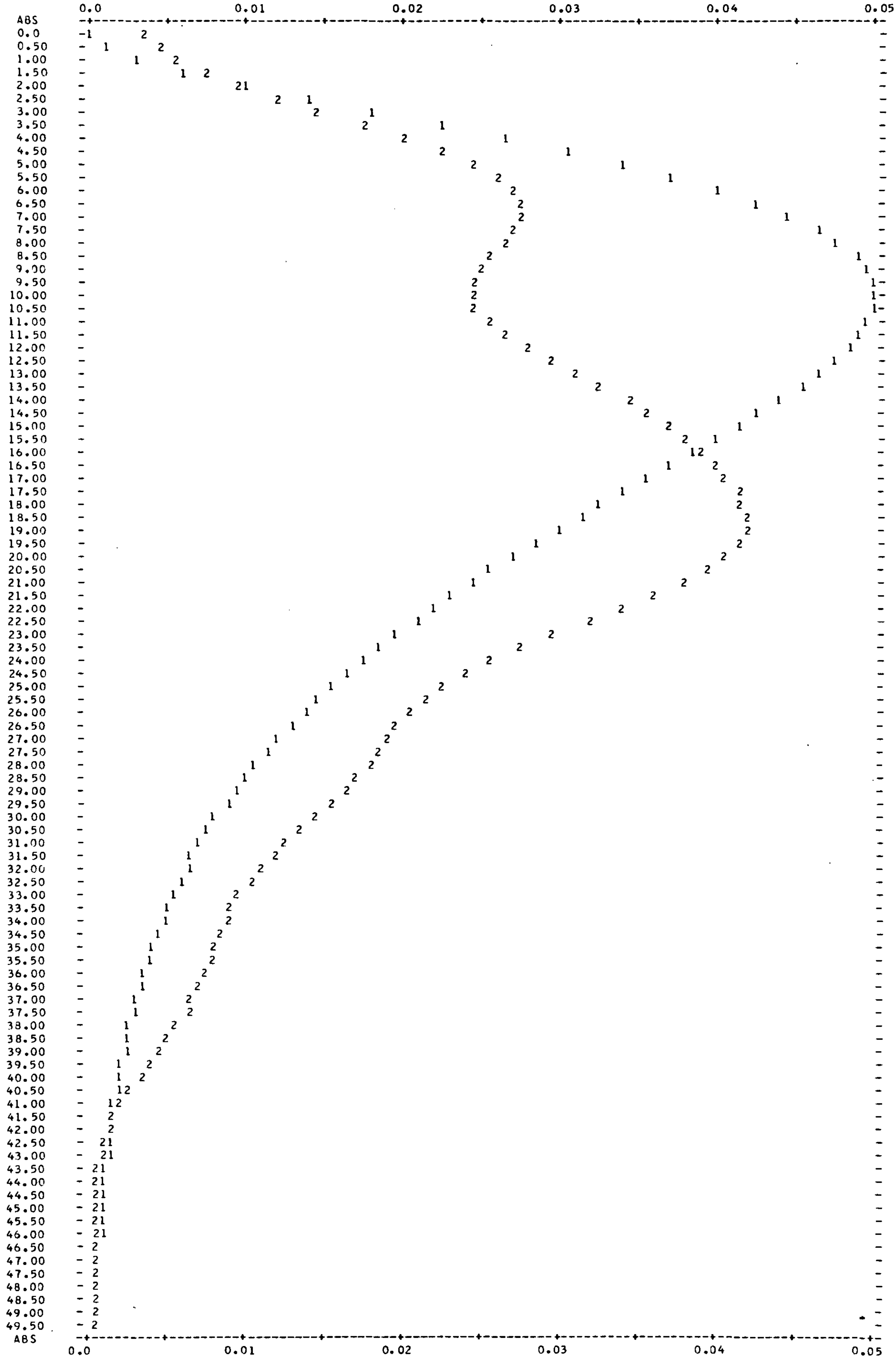
GRAPH

PLOTS OF TRUE AND ESTIMATED DENSITY FUNCTIONS
1 DENOTES TRUE G(3,.2), 2 DENOTES ESTIMATED
K(Y) = (1/(2*PI))*(SIN(Y/2)/(Y/2))**2
N = 25 S = 1.00



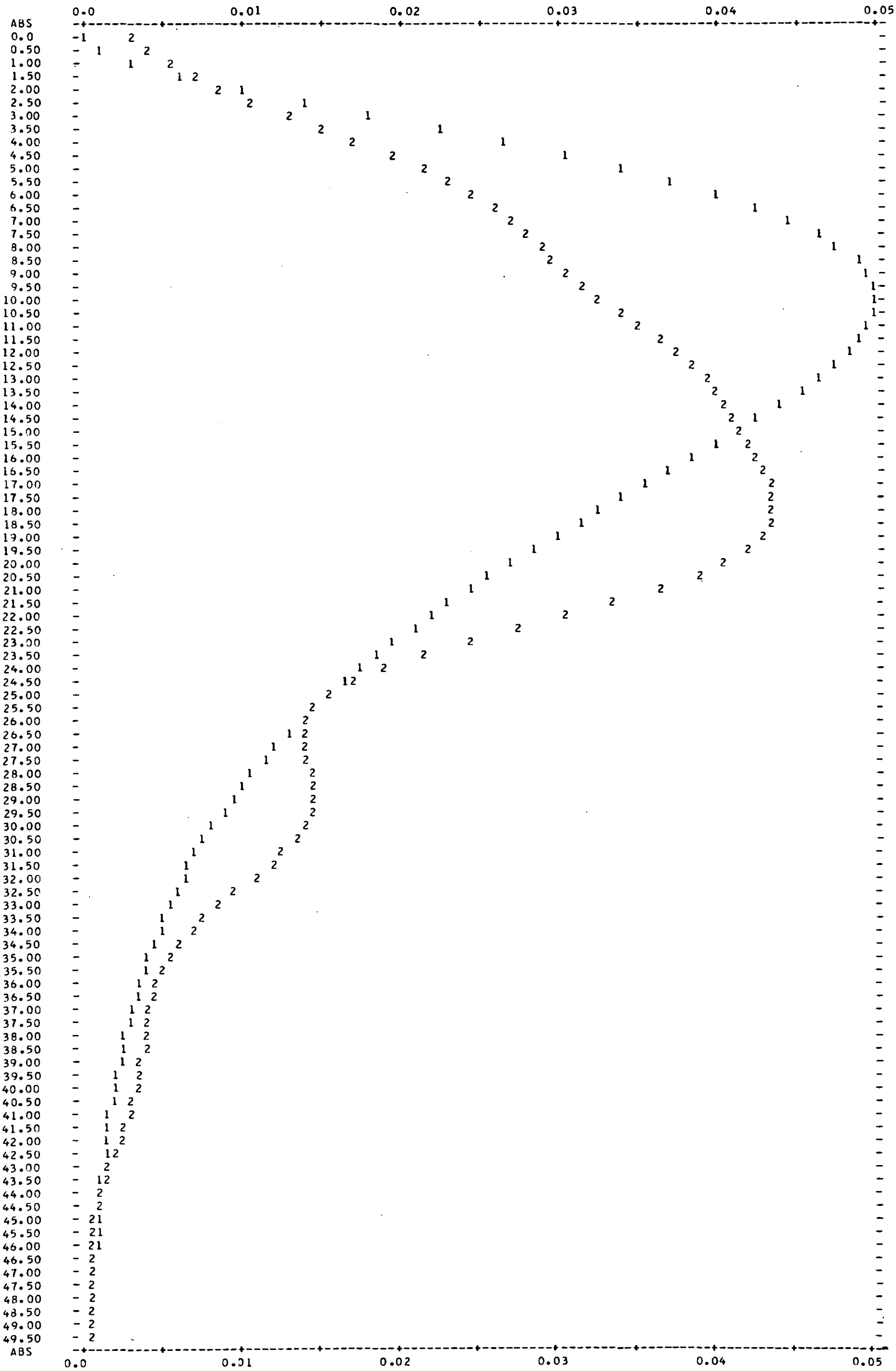
GRAPH

PLOTS OF TRUE AND ESTIMATED DENSITY FUNCTIONS
1 DENOTES TRUE G(3.,.2), 2 DENOTES ESTIMATED
K(Y) = (1/(2*PI))*(SIN(Y/2)/(Y/2))**2
N = 50 S = 1.00



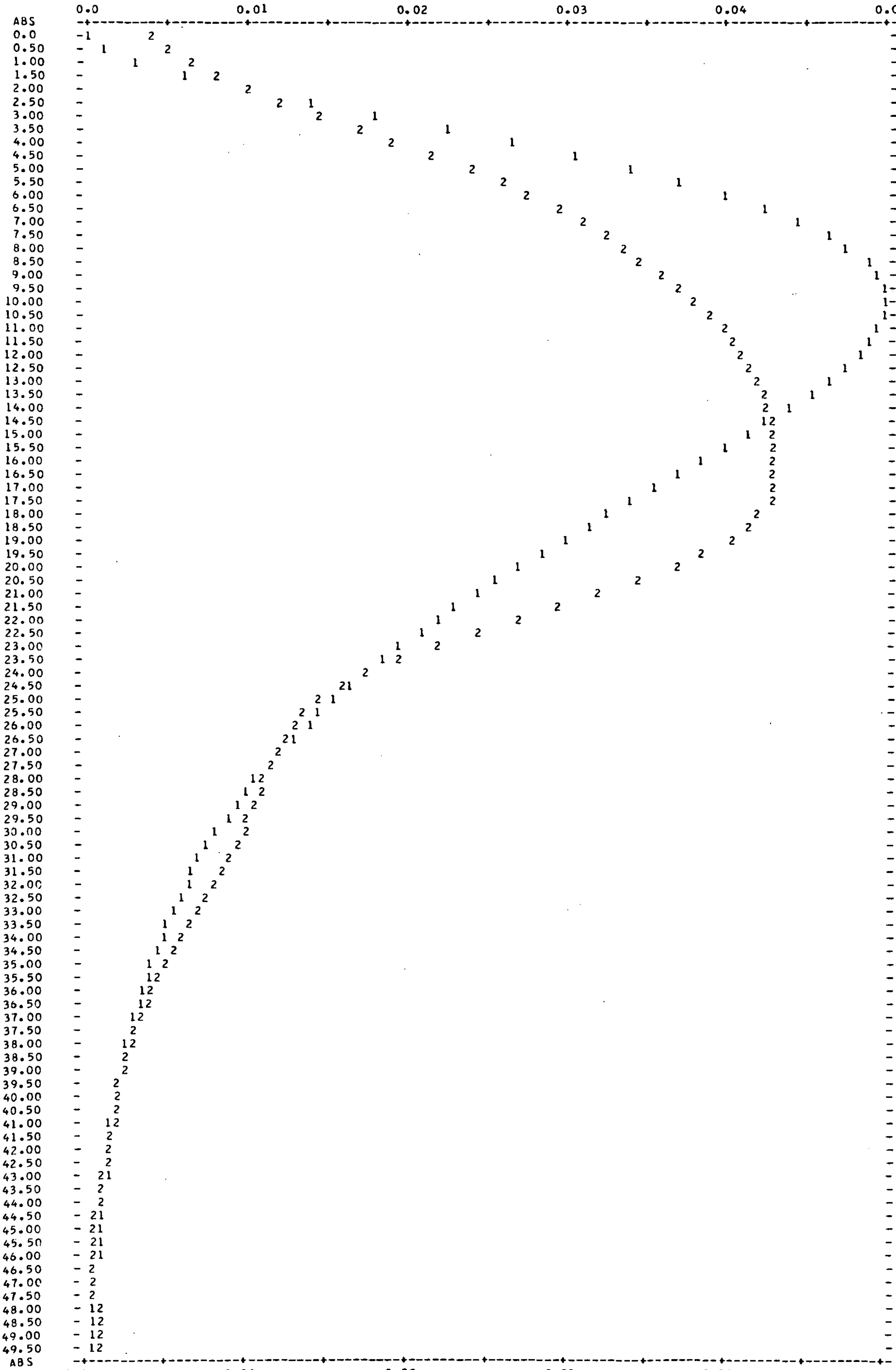
GRAPH

PLOTS OF TRUE AND ESTIMATED DENSITY FUNCTIONS
1 DENOTES TRUE G(3,.2), 2 DENOTES ESTIMATED
K(Y) = (1/(2*PI))*(SIN(Y/2)/(Y/2))**2
N = 100 S = 1.00



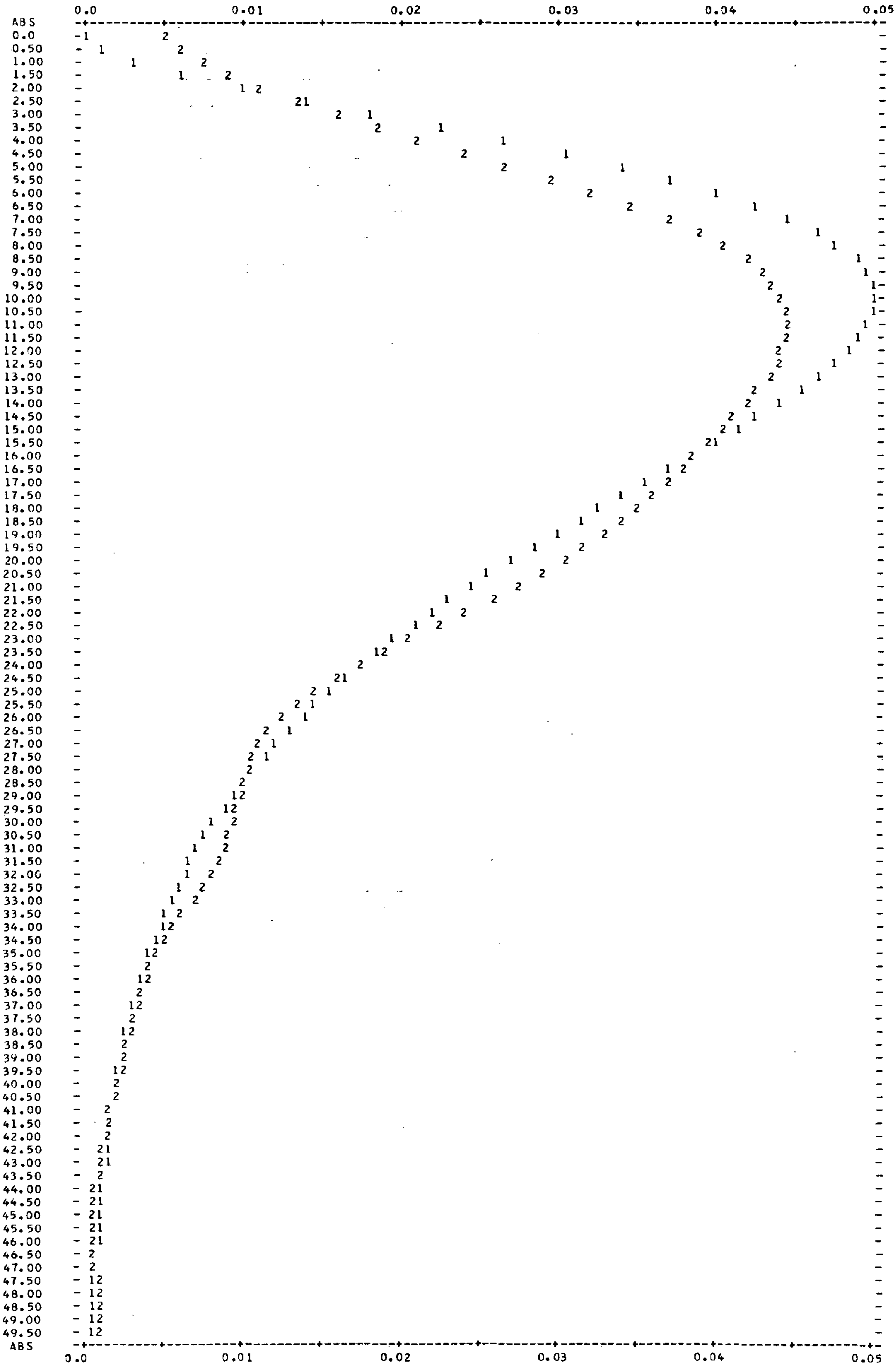
GRAPH

PLOTS OF TRUE AND ESTIMATED DENSITY FUNCTIONS
1 DENOTES TRUE G(3,.2), 2 DENOTES ESTIMATED
K(Y) = (1/(2*PI))*(SIN(Y/2)/(Y/2))**2
N = 200 S = 1.00



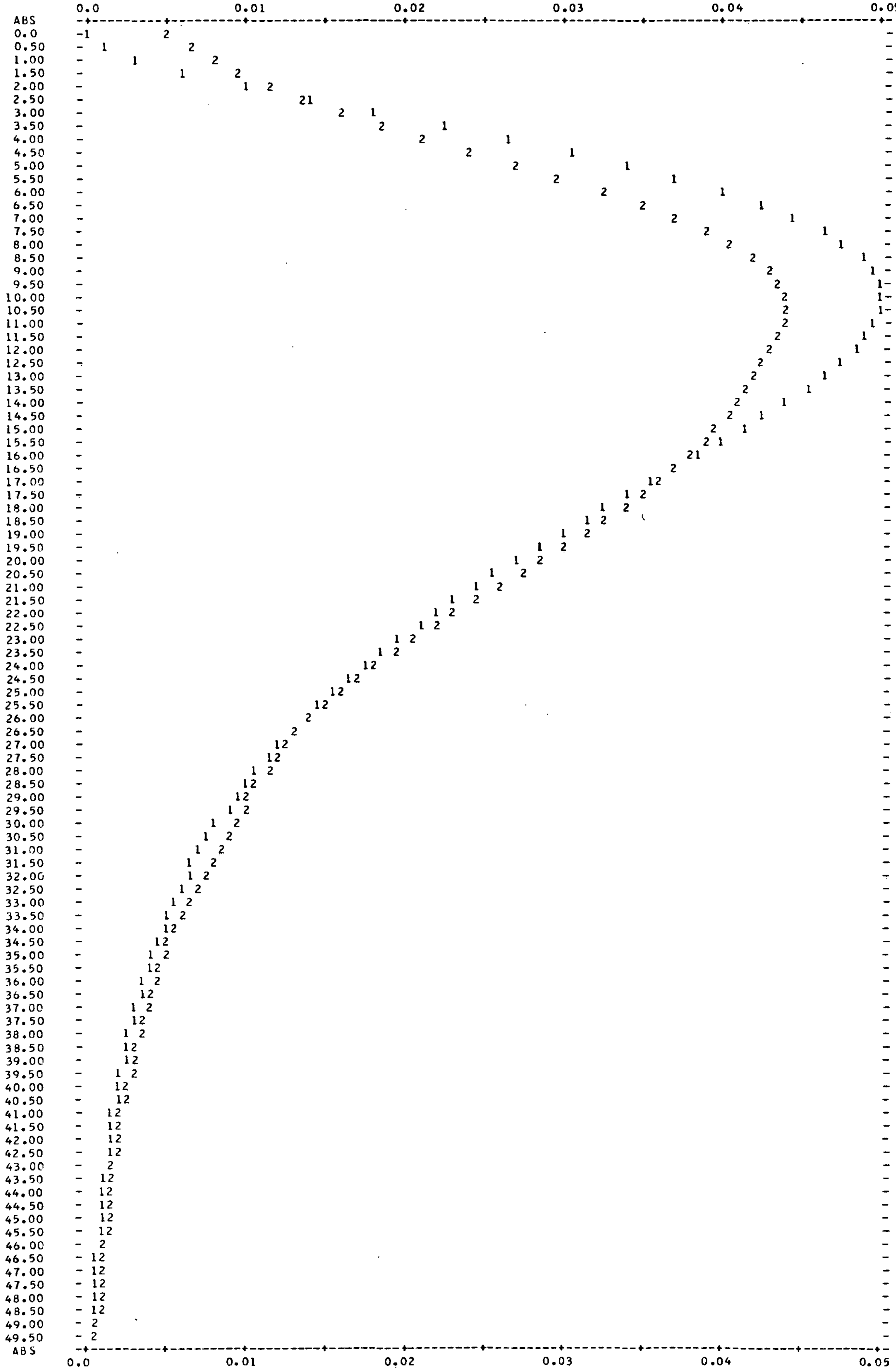
GRAPH

PLOTS OF TRUE AND ESTIMATED DENSITY FUNCTIONS
1 DENOTES TRUE G(3,.2), 2 DENOTES ESTIMATED
K(Y) = (1/(2*PI))*(SIN(Y/2)/(Y/2))**2
N = 500 S = 1.00



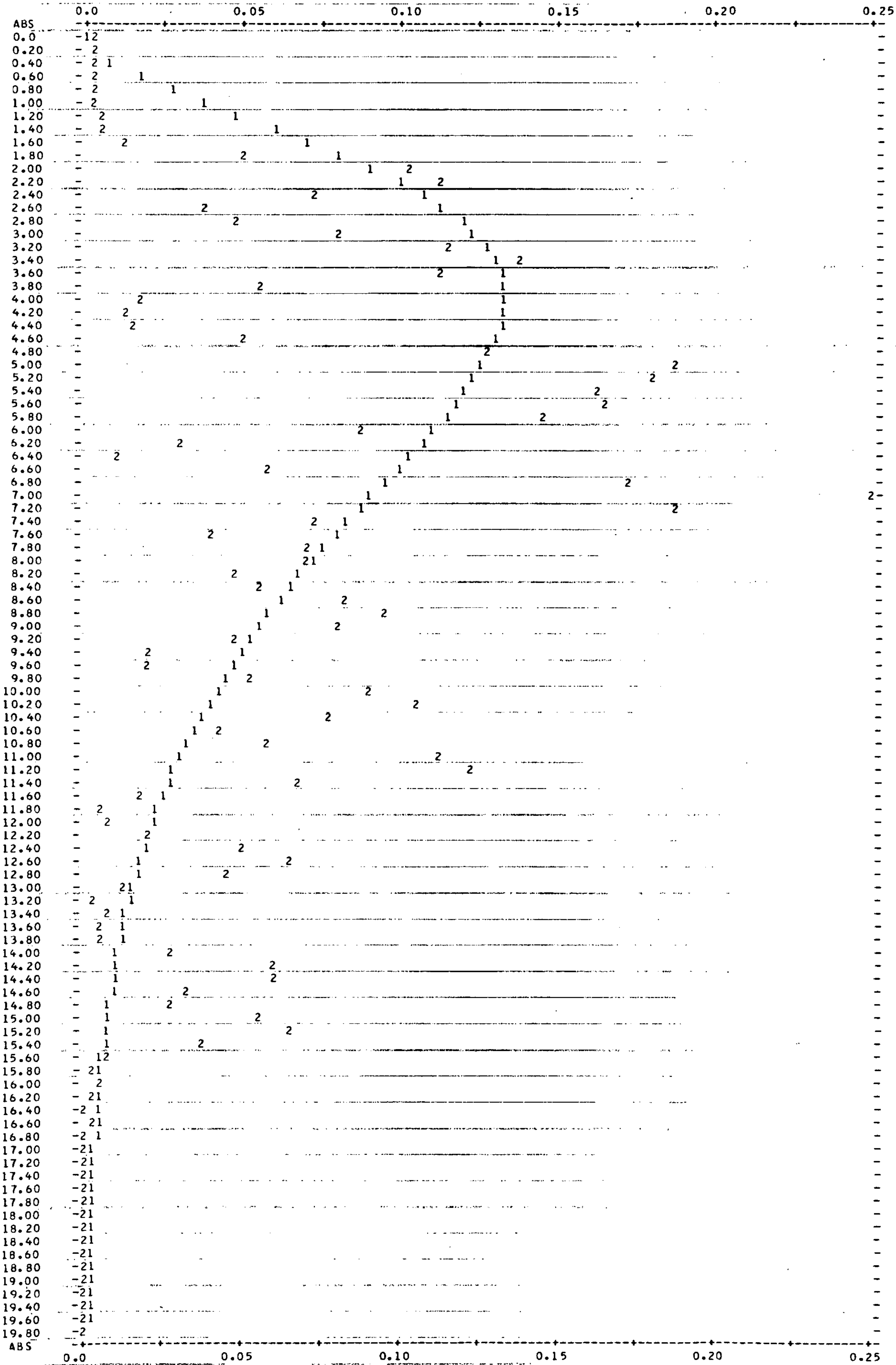
GRAPH

PLOTS OF TRUE AND ESTIMATED DENSITY FUNCTIONS
1 DENOTES TRUE G(3,.2), 2 DENOTES ESTIMATED
K(Y) = (1/(2*PI))*(SIN(Y/2)/(Y/2))**2
N = 1000 S = 1.00



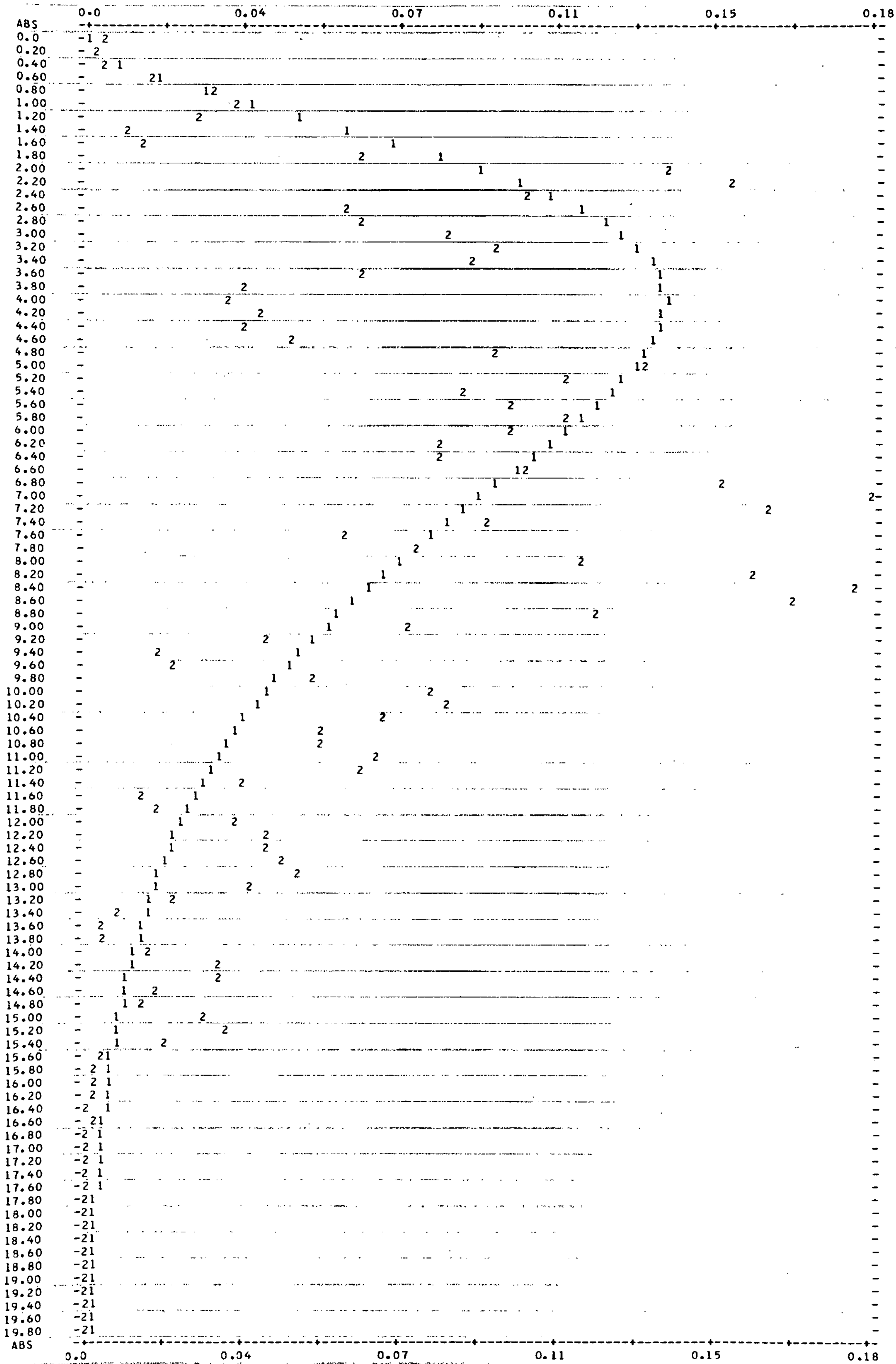
GRAPH

PLOTS OF TRUE AND ESTIMATED DENSITY FUNCTIONS
1 DENOTES TRUE G(3,.5), 2 DENOTES ESTIMATED
K(Y) = (1/(2*PI))*(SIN(Y/2)/(Y/2))**2
N = 25 S = 0.10



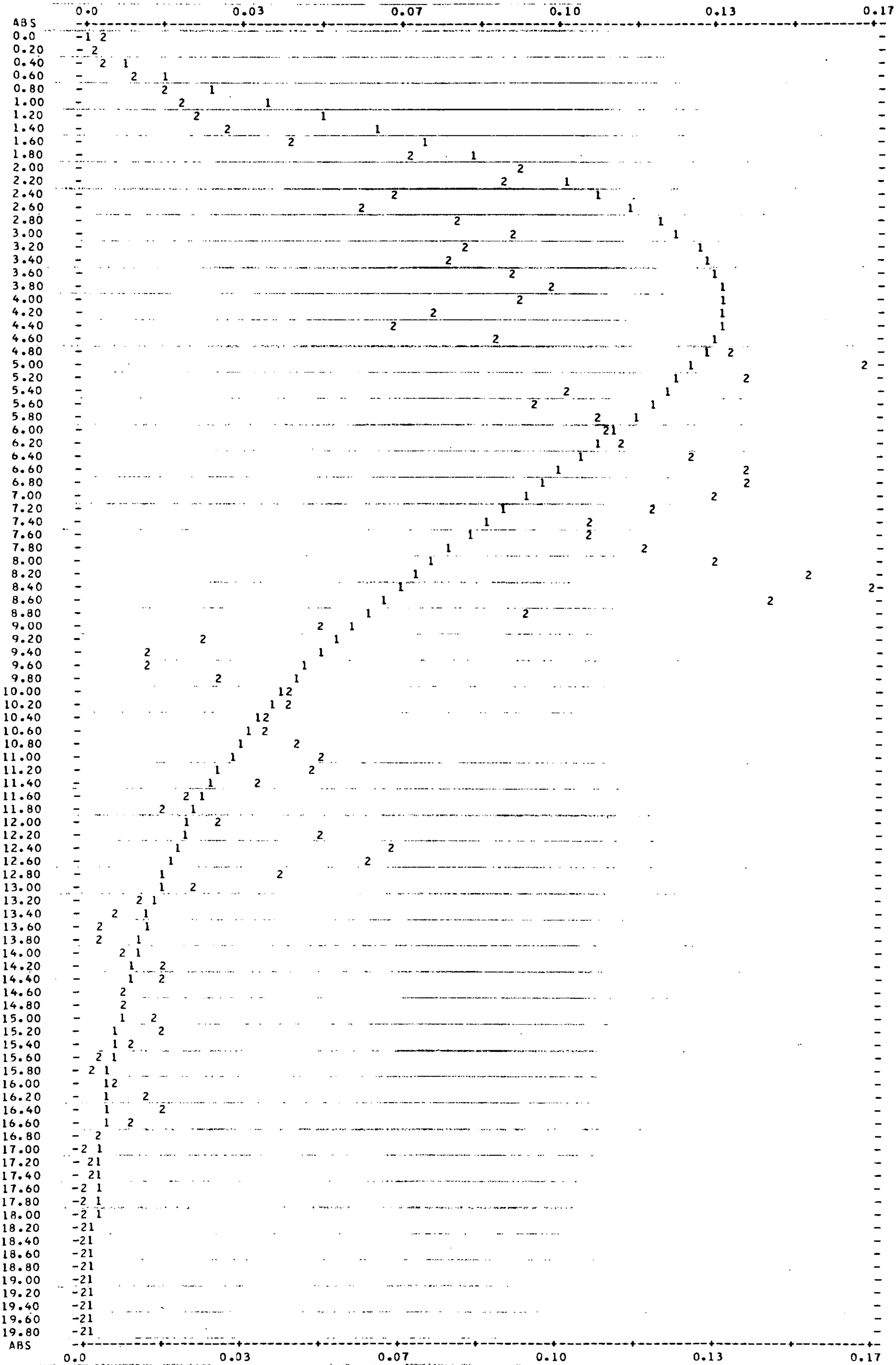
GRAPH

PLOTS OF TRUE AND ESTIMATED DENSITY FUNCTIONS
1 DENOTES TRUE G(3,.5), 2 DENOTES ESTIMATED
K(Y) = (1/(2*PI))*(SIN(Y/2)/(Y/2))**2
N = 50 S = 0.10



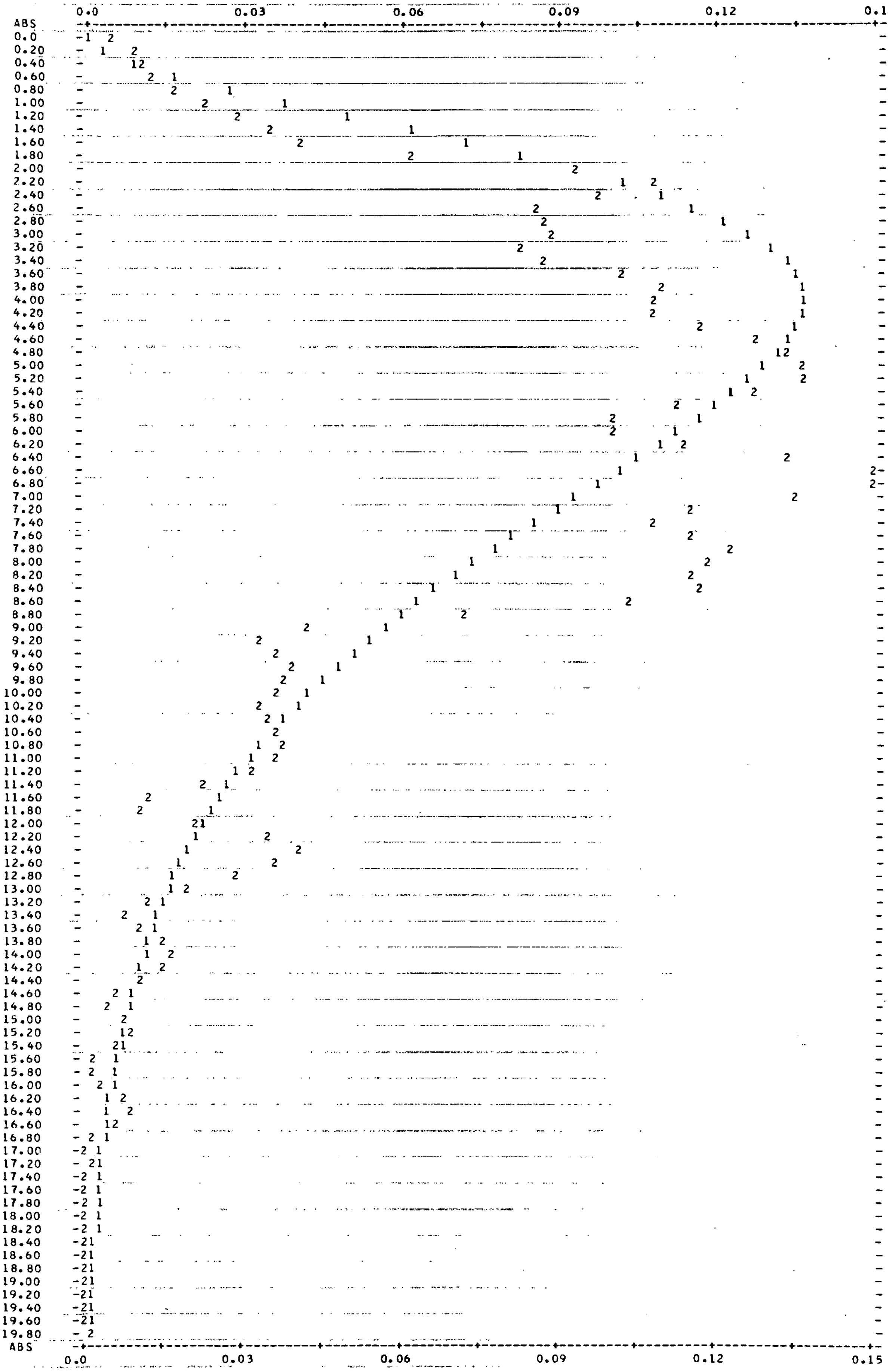
GRAPH

PLOTS OF TRUE AND ESTIMATED DENSITY FUNCTIONS
1 DENOTES TRUE G(3,.5), 2 DENOTES ESTIMATED
K(Y) = (1/(2*PI))*(SIN(Y/2)/(Y/2))**2
N = 100 S = 0.10



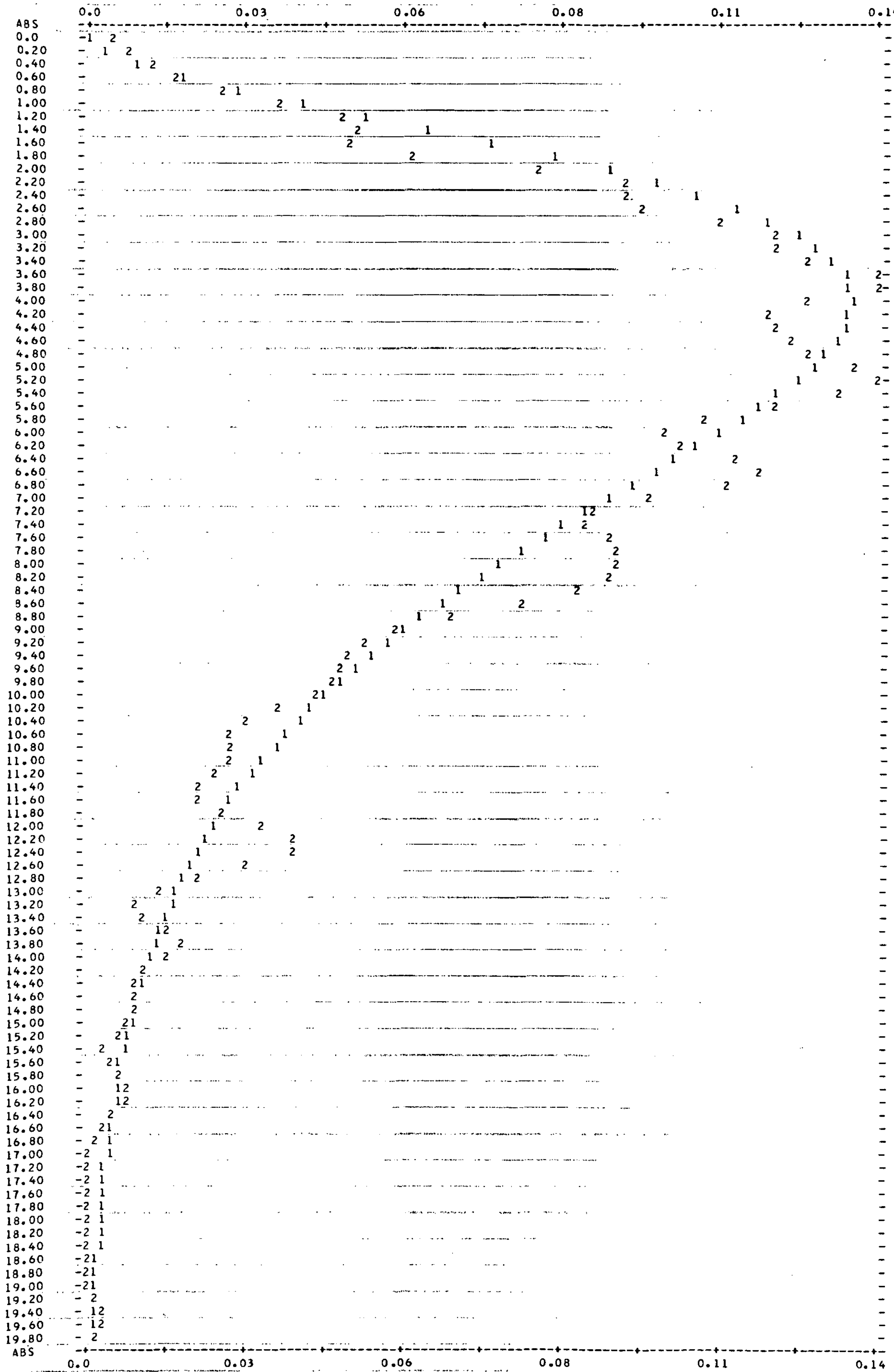
GRAPH

PLOTS OF TRUE AND ESTIMATED DENSITY FUNCTIONS
1 DENOTES TRUE G(3,.5), 2 DENOTES ESTIMATED
K(Y) = (1/(2*PI))*(SIN(Y/2)/(Y/2))**2
N = 200 S = 0.10



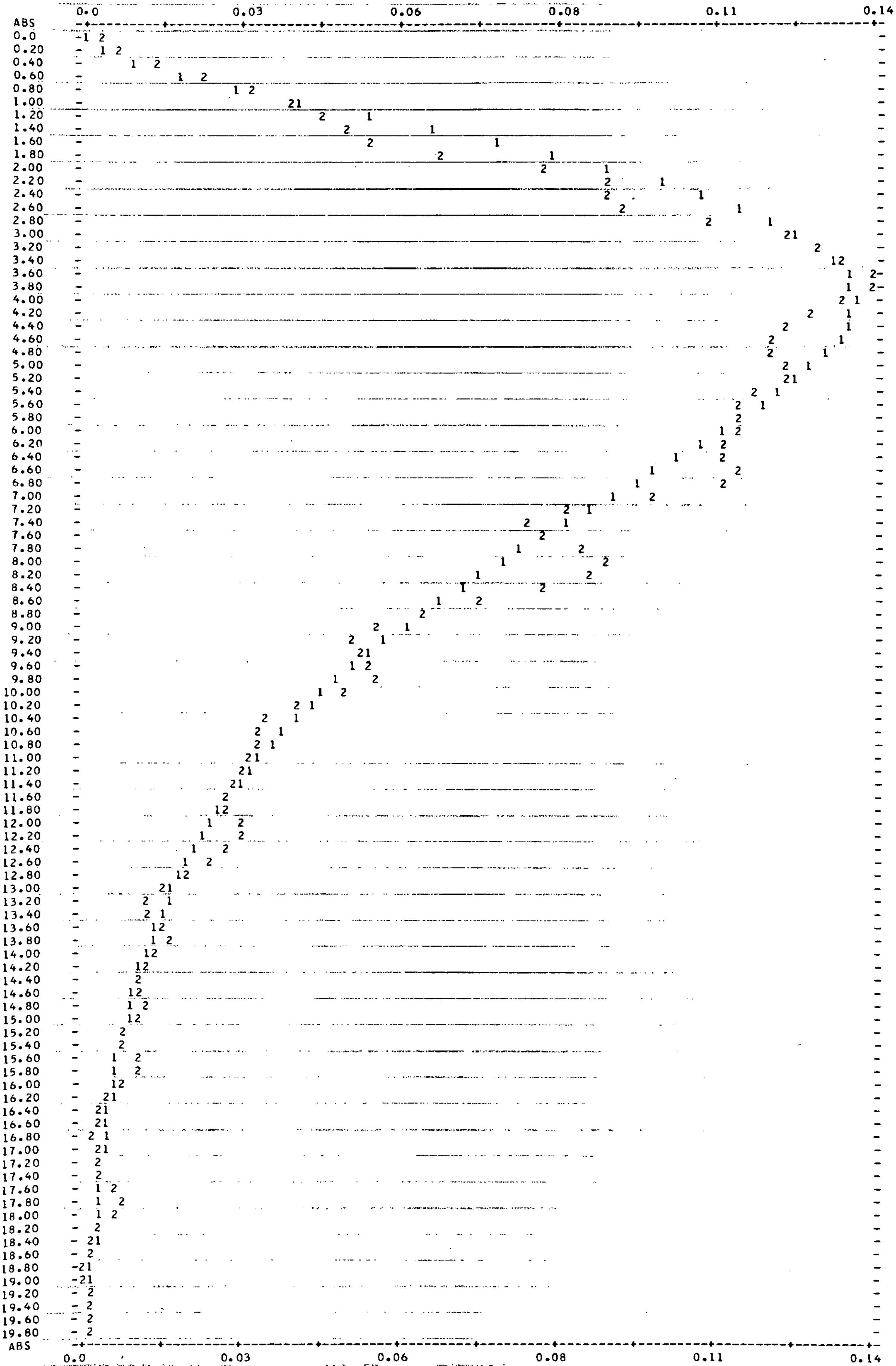
GRAPH

PLOTS OF TRUE AND ESTIMATED DENSITY FUNCTIONS
1 DENOTES TRUE G(3,.5), 2 DENOTES ESTIMATED
K(Y) = (1/(2*PI))*(SIN(Y/2)/(Y/2))**2
N =   500   S = 0.10



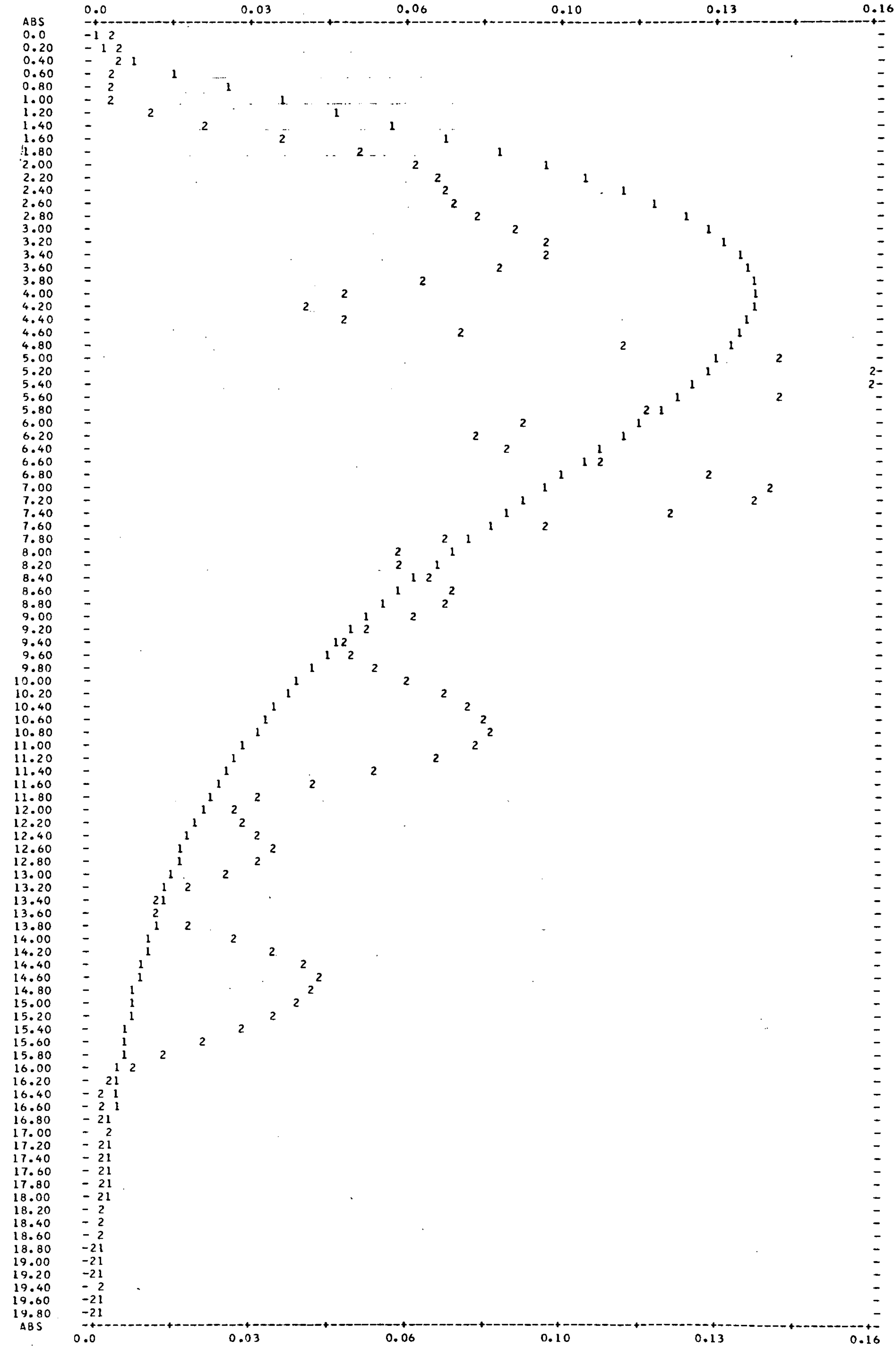
GRAPH

PLOTS OF TRUE AND ESTIMATED DENSITY FUNCTIONS
1 DENOTES TRUE G(3,.5), 2 DENOTES ESTIMATED
K(Y) = (1/(2*PI))*(SIN(Y/2)/(Y/2))**2
N = 1000 S = 0.10



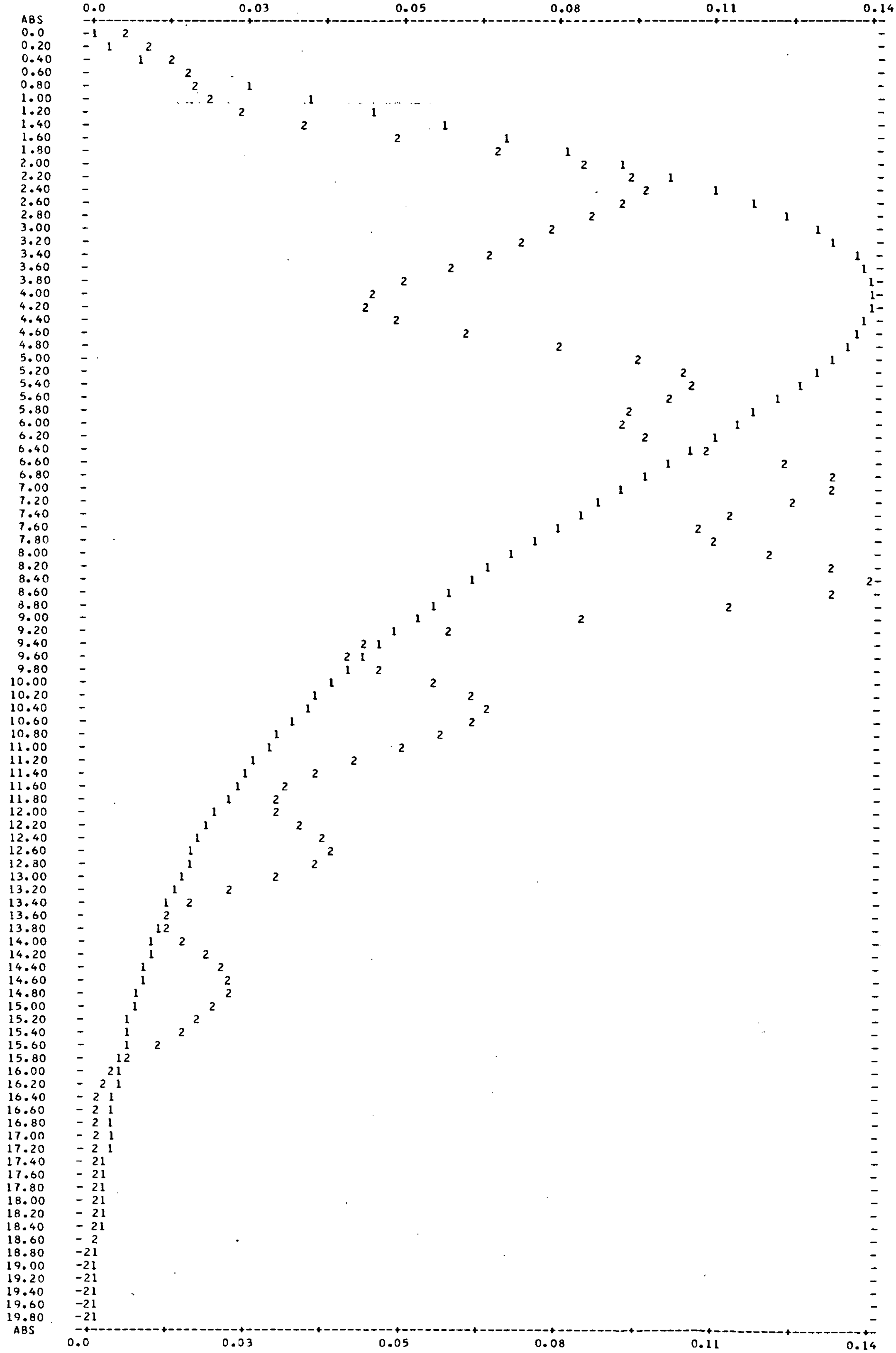
GRAPH

PLOTS OF TRUE AND ESTIMATED DENSITY FUNCTIONS
1 DENOTES TRUE G(3,.5), 2 DENOTES ESTIMATED
K(Y) = (1/(2*PI))*(SIN(Y/2)/(Y/2))**2
N = 25 S = 0.20



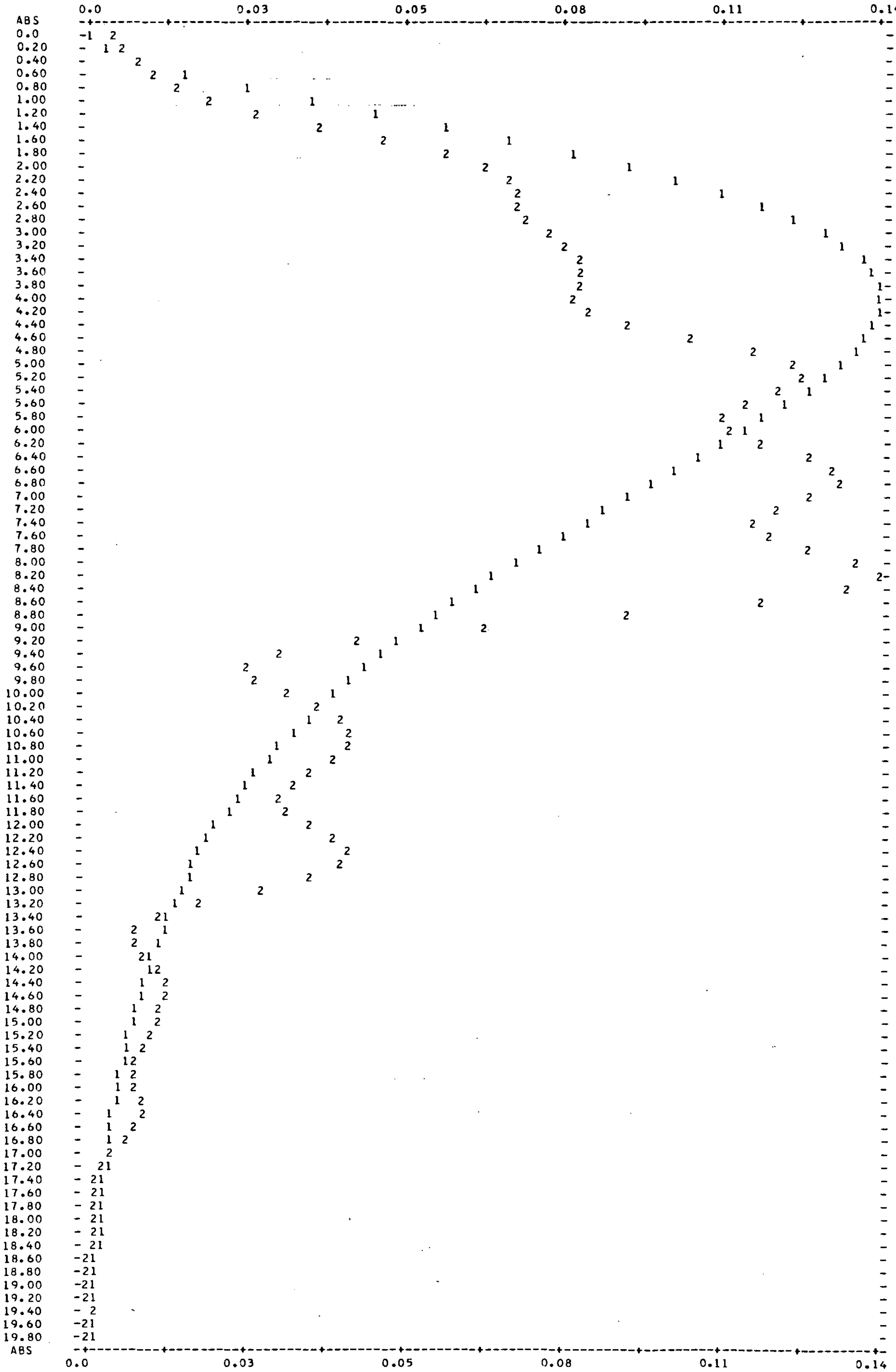
GRAPH

PLOTS OF TRUE AND ESTIMATED DENSITY FUNCTIONS
1 DENOTES TRUE G(3,.5), 2 DENOTES ESTIMATED
K(Y) = (1/(2*PI))*(SIN(Y/2)/(Y/2))**2
N = 50 S = 0.20



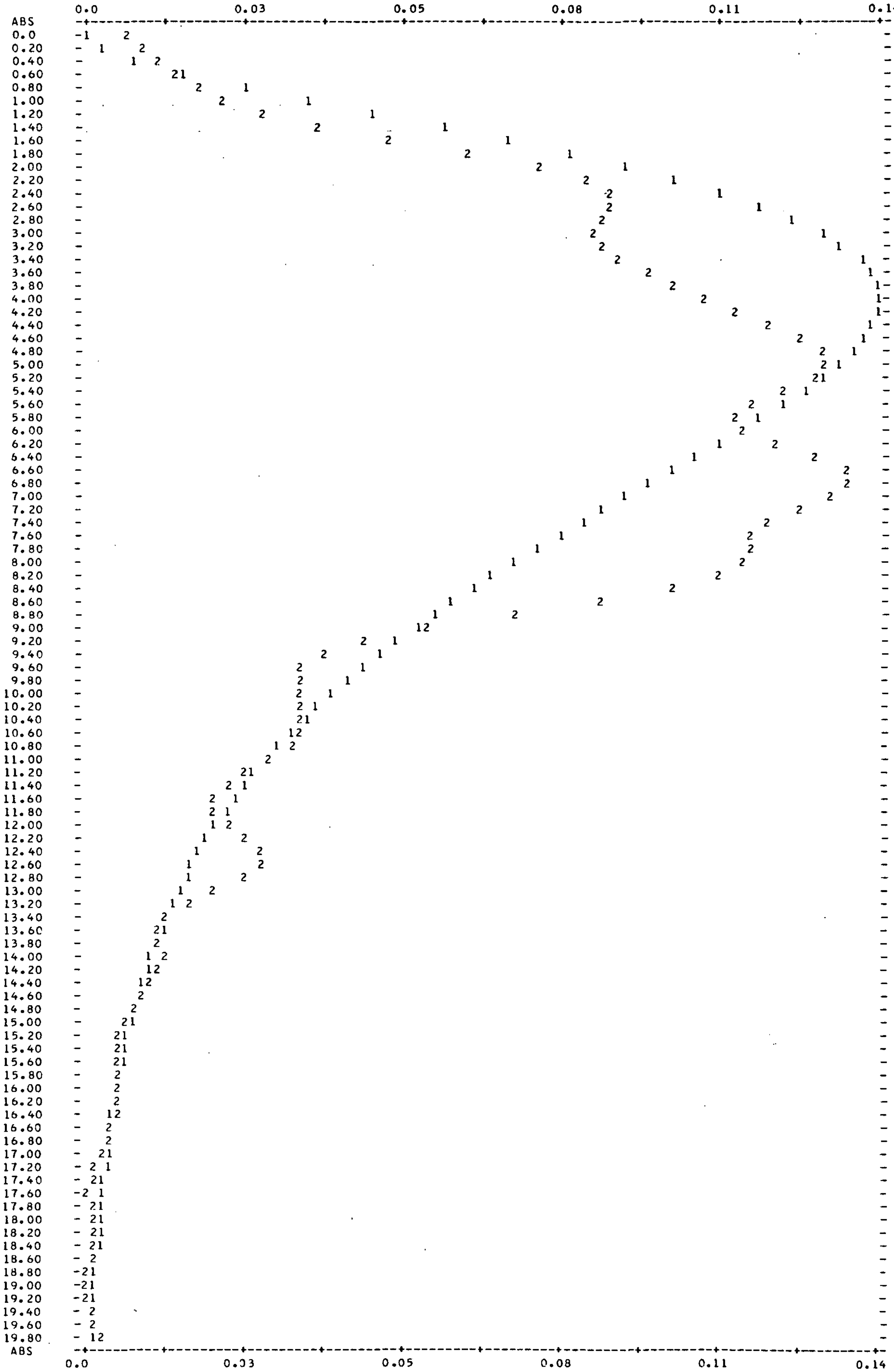
GRAPH

PLOTS OF TRUE AND ESTIMATED DENSITY FUNCTIONS
1 DENOTES TRUE G(3,.5), 2 DENOTES ESTIMATED
K(Y) = (1/(2*PI))*(SIN(Y/2)/(Y/2))**2
N = 100 S = 0.20



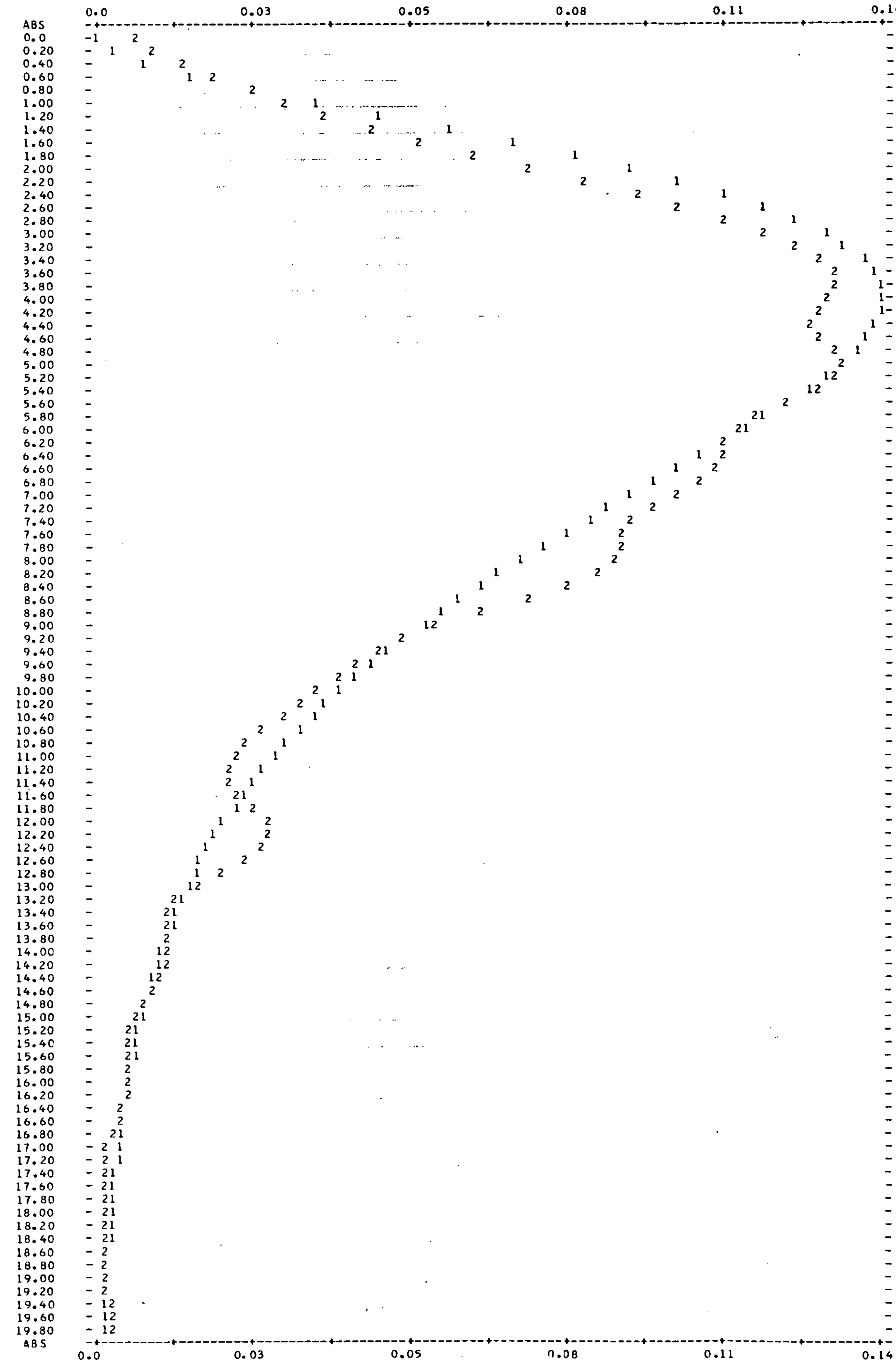
GRAPH

PLOTS OF TRUE AND ESTIMATED DENSITY FUNCTIONS
1 DENOTES TRUE G(3,.5), 2 DENOTES ESTIMATED
K(Y) = (1/(2*PI))*(SIN(Y/2)/(Y/2))**2
N = 200 S = 0.20



GRAPH

PLOTS OF TRUE AND ESTIMATED DENSITY FUNCTIONS
1 DENOTES TRUE G(3,.5), 2 DENOTES ESTIMATED
K(Y) = (1/(2*PI))*(SIN(Y/2)/(Y/2))**2
N = 500 S = 0.20



GRAPH

PLOTS OF TRUE AND ESTIMATED DENSITY FUNCTIONS
1 DENOTES TRUE G(3,.5), 2 DENOTES ESTIMATED
K(Y) = (1/(2*PI))*(SIN(Y/2)/(Y/2))**2
N = 1000 S = 0.20